PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C12N 15/29, 15/63, 15/67
C12N 15/82, A01H 4/00, 5/00

A1

(11) International Publication Number: **WO 91/09948**

(43) International Publication Date: 11 July 1991 (11.07.91)

(21) International Application Number: PCT/US91/00073

(22) International Filing Date: 4 January 1991 (04.01.91)

(30) Priority data:
461,490 5 January 1990 (05.01.90) US

(71) Applicant: CORNELL RESEARCH FOUNDATION, INC. [US/US]; 20 Thornwood Drive, Ithaca, NY 14850 (US).

(72) Inventors: WU, Ray ; 111 Christopher Circle, Ithaca, NY 14850 (US). McELROY, David ; 96 Lowercreek Road, Etna, NY 13062 (US).

(74) Agent: YAHWAK, George, M.; Yahwak & Associates, 25 Skytop Drive, Trumbull, CN 06611 (US).

(81) Designated States: AT (European patent), AU, BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), GR (European patent), HU, IT (European patent), JP, KR, LU (European patent), NL (European patent), SE (European patent), SU.

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: RICE ACTIN GENE AND PROMOTER

(57) Abstract

The present invention describes the genomic nucleotide sequence, and isolation of a strong intron-dependent promoter from rice actin genes.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**ES**	Spain	**MG**	Madagascar
AU	Australia	**FI**	Finland	**ML**	Mali
BB	Barbados	**FR**	France	**MN**	Mongolia
BE	Belgium	**GA**	Gabon	**MR**	Mauritania
BF	Burkina Faso	**GB**	United Kingdom	**MW**	Malawi
BG	Bulgaria	**GN**	Guinea	**NL**	Netherlands
BJ	Benin	**GR**	Greece	**NO**	Norway
BR	Brazil	**HU**	Hungary	**PL**	Poland
CA	Canada	**IT**	Italy	**RO**	Romania
CF	Central African Republic	**JP**	Japan	**SD**	Sudan
CG	Congo	**KP**	Democratic People's Republic of Korea	**SE**	Sweden
CH	Switzerland			**SN**	Senegal
CI	Côte d'Ivoire	**KR**	Republic of Korea	**SU**	Soviet Union
CM	Cameroon	**LI**	Liechtenstein	**TD**	Chad
CS	Czechoslovakia	**LK**	Sri Lanka	**TG**	Togo
DE	Germany	**LU**	Luxembourg	**US**	United States of America
DK	Denmark	**MC**	Monaco		

RICE ACTIN GENE AND PROMOTER

Significant progress has recently been made in the application of tissue culture and gene transfer techniques to previously recalcitrant monocotyledonous plants, such as rice. Rice protoplasts have been shown to transiently express a foreign gene after transformation by electroporation, and stable genomic integration of foreign DNA, following polyethylene-glycol-mediated transformation of rice cells, has been previously demonstrated.

One of the major limitations in rice transformation technology has been the lack of an efficient promoter for high level constitutive expression of foreign genes in transgenic plants. The cauliflower mosaic virus (CaMV) 35S promoter has been widely used for this purpose in a number of plant transformation systems, however, the CaMV 35S promoter has shown low activity in transforming rice cells, and recent reports suggest that the pattern of CaMV 35S promoter activity in transgenic plants may not be constitutive. Work in our laboratory suggests that the 5' region of the maize Adh1 gene containing the promoter, 5'-coding exon and 5'-intron, is 20 to 30 times more active than the CaMV 35S promoter for transient gene expression in transformed rice protoplasts and calli, however, the maximum activity of AdH1 promoter in transformed rice protoplasts requires anaerobic stress and its activity is not constitutive in all transformed rice tissues. As a result, a number of alternative

promoter elements for rice transformation are currently under investigation.

It is, therefore, the purpose of the present invention to describe an efficient promoter for rice transformation, and in doing so have based such a promoter upon a rice actin gene that displays a constitutive spatial and temporal pattern of transcript abundance.

Cytoplasmic actin is a fundamental component of the eukaryotic cell cytoskeleton. In higher-plant cells a number of cellular process, such as cytoplasmic streaming, extension growth and cell division, are believed to involve cytoskeletal actin protein. Actin has been found to be encoded by gene families in all higher plants studied to date. A unique feature of these plant actin genes, in contrast to the situation for animal actin genes, has been their conservation of gene structure. All of the studied plant actin genes are reported to consist of four coding exons of conserved length, separated by three introns of variable length. In each of these cases the coding regions of the plant actin genes have been deduced following a comparison of their potential translations products with that of previously published animal actin protein sequences. However, in a number of animal actin genes there exists a 5' transcribed, but untranslated, sequence (5'-noncoding exon), separated by an intron (5'-intron) from the exon containing the first translated codon (5'-coding exon). Although it has been suggested that a 5'-noncoding exon may exist in plant actin genes, such an exon would

fail to be detected in any comparison of the coding region of plant and animal actin genes.

In rice, there are at least 8 actin-like sequences per haploid genome, of which four have been isolated and shown to differ from each other in the tissue and stage specific abundance of their respective transcripts. One rice actin gene, RAc1, was found to encode a transcipt that is relatively abundant in all rice tissues, and at all developmental stages examined. In view of the high level constitutive abundance of the rice RAc1 transcript the present invention describes the isolation and partial sequencing of a genomic clone containing the rice RAc1 5' flanking and 5'-transcribed, non-translated, regions. From such clones, a number of RAc1-GUS (GUS stands for ß-glucuromidase gene) fusions plasmids were constructed and used in transient expression assays of transformed rice protoplasts. The results from these assays suggest that the regulatory element(s) necessary for maximal RAc1 promoter activity in transformed rice protoplasts are located within a region 1.3 kb upstream of the RAc1 translation initiation codon. The rice RAc1 promoter is 5 times more active than the maize Adh1 promoter in transformed rice cells, suggesting that the RAc1 5' region contains an efficient promoter for rice transformation.

As a first step towards a characterization of the actin gene in rice, several clones from a rice genomic library representing four unique actin sequences were isolated. A rice (*Oryza sativa* variety IR26) genomic library in lambda-EMBL4 was screened

with a heterologous actin probe. Fourteen independent clones were isolated and subcloned into pUC13. By mapping restriction sited and carrying out cross-hybridization studies, four different classes of clones were identified and designated RAc1, RAc2, RAc3 and RAc7.

Sequencing of the subcloned actin genes was carried out by the dideoxynucleotide chain termination method and computer analysis of the resulting sequences was done. The nucleotide sequences of the coding regions of the four actin genes appear in the EMBL Gene Bank and DDB7 Nucleotide Sequence Databases under the accession numbers X15865 RAc1; X15864 RAc2; X15862 RAc3; and X15863 RAc7.

Prior to identifying those sequences which regulate RAc1 expression, its complete genomic structure was determined. By characterizing an RAc1 cDNA clone (EMBL data bank accession number X16280) and 5' end mapping the RAc1 transcript, the structure of the RAc1 gene was determined and the position of a 5'-noncoding exon in its genomic sequence was identified and located. This represents the first complete structural characterization of plant actin gene and is one of few reported cases of a 5'-noncoding exon in a plant gene.

A complete understanding of all aspects of the present invention can be better understood from reference to the accompanying figures (and examples) in which:

Figure 1a is a restriction map of the pcRAc1.3 cDNA insert, according to the present invention;

Figure 1b is a restriction map and proposed structure of the RAc1 genomic clone according the the present invention;

Figure 2a is a restriction map of the 15.1 Kb lambda-RAc1 insert, according to the present invention;

Figure 2b is a restriction map of the pRAc 15'.H3 insert according to the present invention;

Figure 3 depicts maps of constructs according to the present invention.

The identification of the complete genomic structure was accomplished utilizing the protocols described in the following examples:

Example 1

a. Screening the lambda-gt11 c DNA library:

A rice (*Oryza sativa* IR36) lambda-gt11 cDNA library, prepared from six-day-old etiolated leaf tissue was plated and lifted onto nylon filters. Primary and secondary screenings for rice actin cDNA clones were carried out by hybridization with a 260 bp, ^{32}P-dATP labeled BamH1-HindIII restriction fragment from the rice actin genomic clone pRAc2. Tertiary screening was carried out using rice actin gene- specific probes. RAc1-like cDNA clones were identified using a 900-bp BamH1-HindIII restriction fragment probe from the 3' end of the rice actin genomic clone pRAc1.

b. Restriction mapping and nucleotide sequencing:

Actin positive clones from tertiary screening of the lambda-gt11 cDNA library were further characterized by restriction mapping. Restriction fragments from the cDNA insert of the RAc1 positive clone pcRAc1.3 were subcloned into pBS-KS vectors for nucleotide sequencing. Double-stranded DNA sequencing using ^{35}S-dATP was carried out using a T7 polymerase sequencing kit following the manufacturer's (Pharmacia) modification of the Sanger *et al* dideoxynucleotide sequencing procedure. The 5' and 3' ends of the pcRAc1.3 cDNA insert were further sequenced using two synthetic oligonucleotides; 5' -AAGCGGCCTCTACTACGC-3' and 5'-

GAAGCATTTCCTGRGCACAAT-3' respectively and subsequently the sequence data were analyzed.

c. 5'-end mapping of RAc1 transcript by northern blot analysis and primer extension:

RNA isolation from seven-day-old rice shoots and Northern blotting were performed. 10 µg of total RNA samples were used in formaldehyde agarose gel electrophoresis and northern analysis was carried out under stringent hybridization conditioned (50% formamide, 50°C). Restriction fragment probes for northern hybridization were isolated from the 5'-untranslated end of the rice actin genomic clone pRAc1 and subcloned into pBS-KS vectors.

The primer extension analysis was performed using 1 µg of a synthetic oligonucleotide primer 5'-CTTCTACCGGCGGCGGC-3' which was annealed to 25 µg of total RNA from seven-day-old rice shoots.

The cDNA library made with mRNA from six-day-old etiolated rice shoots according to Example 1 provided the cDNA clones corresponding to the multiple members of the rice actin gene family. Primary and secondary screening with a 260 bp *Bam* H1-*Hind* III restriction fragment - a fragment previously known to cross-hybridize with many members of the rice actin gene family - from the rice actin gene RAc2, because of the homology found among isolated rice actin genomic sequenced. Thirty four clones were identified which strongly hybridized to the non-specific RAc2 actin probe.

Tertiary screening of the 34 actin-positive cloned was carried out using probes previously determined to be actin gene specific. Seven RAc1-like cDNA clones were identified which hybridized specifically to a 900 bp BamH1-HindIII restriction fragment from pRAc1. The nature of these clones was confirmed by restriction mapping and Southern blotting. The EcoRI insert from the lambda-gt 11 clone lambda- RAc1.3 was subcloned into a pBS-KS vector to produce the plasmid pcRAc1.3.

Using similar procedures, nine RAc3-like cDNA clones and eight RAc7-like cDNA clones were identified. No cDNA clones were identified which cross-hybridized with a RAc2 gene specific probe. The remaining ten actin-positive clones failed to cross-hybridize with any of the previously characterized actin gene-specific probes. These were classified into five groups on the basis of restriction mapping and cross-hybridization analysis. The isolation of cDNA clones representing eight distinct actin

transcripts confirms that the actin gene family in rice composed of at least eight unique members.

The insert from pcRA1.3 was further subcloned into pBS-KS plasmids. Double stranded DNA was prepared for sequencing. The restriction enzyme map of the pcRAC1.3 insert and the sequencing strategy for determining its base are summarized in Fig. 1a.

More specifically, Fig. 1 describes the determination of the RAc1 gene structure by alignment of the pcRAc1.3 cDNA insert with the pRAc1 genomic clone a: restriction map of the pcRAc1 cDNA insert with the strategy used to determine its base sequence. The horizontal arrows indicate the direction of sequencing and their length indicate the actual size of the sequence obtained. Horizontal arrows preceded with a sinusoidal wave indicate the use of synthetic oligonucleotide primers. b: restriction map and proposed structure of the RAc1 genomic clone. Horizontal lines represent introns and boxes represent exons. Open boxes represent the transcribed but untranslated regions of exons, closed boxes represent the translated regions of exons. Restrictions sites: B, BamH1; H, HindIII ; R, EcoRV; S, SstI. Dashed lined between the cDNA and genomic DNA restriction maps indicate the alignment of their homologous exons.

The tabulated sequence which depicts the nucleotide sequence of the rice actin gene RAc1, is depicted in the following sequence of 5510 nucleotides

```
1     GATATCCCTC AGCCGCCTTT CACTATCTTT TTTGCCCGAG
      TCATTGTCAT GTGAACCTTG GCATGTATAA TCGGTGAATT
81    GCGTCGATTT TCCTCTTATA GGTGGGCCAA TGAATCCGTG
      TGATCGCGTC TGATTGGCTA GAGATATGTT TCTTCCTTGT
161   TGGATGTATT TTCATACATA ATCATATGCA TACAAATATT
      TCATTACACT TTATTAGAGG TGGTCAGTAA TAAACCCTAT
241   CACTATGTCT GGTGTTTCAT TTTATTTGCT TTTAAACGAA
      ATTGACTTCT GATTCAATAT TTAAGGATCG CAATCGGCGT
321   GCAGTACTAA TTCTGGTTGA CGGAACTATA CGTAAACTAT
      TCAGCTTCAC TCTATTAGGC ACTACTTGAC ACACCGGAGT
401   ACGTACGAAT ACTCTATCAA TTAGTCTCAG TCTAACTTGT
      TGAGACATGT ACTATAGATT ACTATTGTAC CTTGACTCAC
481   TGTATGTATC ACGTCTAATT GAACTACACA TATATACGCG
      ATATTTTTTA ATAACATTAA AACCTACCTC TATGTCAACA
561   ATGGTGTACG ATAACCACAA GTTTAGGAGG TAAAAAAACA
      TTGCCTTACG CTGAAGTTAC GCCTTAAAAA TAAAGAGTAA
641   ATTTTACTTT GACCACCCTT CAATGTTCAC TTTAGACCGG
      TGGAACGCTC CAGCCGTAAT AGGATTCTGC ACCTCACATG
721   CCTTAGCCGG ATTATATTGC CTGCCCACTT TCTCACTCAT
      ATCTGCAAGA ATGTCTACTC GCTAGAATTA TCGCGATAGT
801   AGCTAGCATA CTCGAGGTCA TTCATATGCT TGAGAAGAGA
      GTCGGGATAG TCCAAAATAA AACAAAGGTA AGATTACCTG
881   GTCAAAAGTG AAAACATCAG TTAAAAGGTG GTATAAGTAA
      AATATCGGTA ATAAAAGGTG GCCCAAAGTG AAATTTACTC
961   TTTTCTACTA TTATAAAAAT TGAGGATGTT TTGTCGGTAC
      TTTGATACGT CATTTTTGTA TGAATTGGTT TTTAAGTTTA
1041  TTCGCGATTT TGGAAATGCA TATCTGTATT TGAGTCGGGT
      TTTAAGTTCG TTTGCTTTTG TAAATACAGA GGGATTTGTA
1121  TAAGAAATAT CTTTAAAAAA ACCCATATGC TAATTTGACA
      TAATTTTTGA GAAAAATATA TATTCAGGCG AATTCTCACA
1201  ATGAACAATA ATAAGATTAA AATAGCTTGC CCCCGTTGCA
      GCGATGGGTA TTTTTTCTAG TAAAATAAAA GATAAACTTA
1281  GACTCAAAAC ATTTACAAAA ACAACCCCTA AAGTCCTAAA
      GCCCAAAGTG CTATGCACGA TCCATAGCAA GCCCAGCCCA
```

1361 ACCCAACCCA ACCCAACCCA CCCCAGTGCA GCCAACTGGC
AAATAGTCTC CACACCCCGG CACTATCACC GTGAGTTGTC
1441 CGCACCACCG CACGTCTCGC AGCCAAAAAA AAAAAAAGAA
AGAAAAAAAA GAAAAAGAAA AAACAGCAGG TGGGTCCGGG
1521 TCGTGGGGGC CGGAAAAGCG AGGAGGATCG CGAGCAGCGA
CGAGGCCGGC CCTCCCTCCG CTTCCAAAGA AACGCCCCCC
1601 ATCGCCACTA TATACATACC CCCCCCTCTC CTCCCATCCC
CCCAACCCTA CCACCACCAC CACCACCACC TCCTCCCCCC
1681 TCGCTGCCGG ACGACGAGCT CCTCCCCCCT CCCCCTCCGC
CGCCGCCGGT AACCACCCCG CCCCTCTCCT CTTTCTTTCT
1761 CCGTTTTTTT TTTCGTCTCG GTCTCGATCT TTGGCCTTGG
TAGTTTGGGT GGGCGAGAGC GGCTTCGTCG CCCAGATCGG
1841 TGCGCGGGAG GGGCGGGATC TCGCGGCTGG CGTCTCCGGG
CGTGAGTCGG CCCGGATCCT CGCGGGGAAT GGGGCTCTCG
1921 GATGTAGATC TTCTTTCTTT CTTCTTTTTG TGGGTAGAAT
TTGAATCCCT CAGCATTGTT CATCGGTAGT TTTTCTTTTC
2001 ATGATTTGTG ACAAATGCAG CCTCGTGCGG AGCTTTTTTG
TAGGTAGAAG ATGGCTGACG CCGAGGATAT CCAGCCCCTC
2081 GTCTGCGATA ATGGAACTGG TATGGTCAAG GTAAGCTGTT
TGGATCTCAG GGTGGTTTCC GTTTACCGAA ATGCTGCATT
2161 TCTTGGTAGC AAAACTGAGG TGGTTTGTGT CAGGCTGGGT
TCGCCGGAGA TGATGCGCCC AGGGCTGTCT TCCCCAGCAT
2241 TGTCGGCCGC CCTCGCCACA CCGGTGTCAT GGTCGGAATG
GGCCAGAAGG ACGCCTACGT CGGCGACGAG GCGCAGTCCA
2321 AGAGGGGTAT CTTGACCCTC AAGTACCCCA TCGAGCATGG
TATCGTCAGC AACTGGGATG ATATGGAGAA GATCTGGCAT
2401 CACACCTTCT ACAACGAGCT CCGTGTGGCC CCGGAGGAGC
ACCCCGTCCT CCTCACCGAG GCTCCTCTCA ACCCCAAGGC
2481 CAATCGTGAG AAGATGACCC AGATCATGTT TGAGACCTTC
AACACCCCTG CTATGTACGT CGCCATCCAG GCCGTCCTCT
2561 CTCTGTATGC CAGTGGTCGT ACCCAAGGTG AGCACATTCG
ACACTGAACT AAAAGGCTGT GAGGATGAAT TTTAATTTTG
2641 ACATTCACAT GTAGATGAGA TTTAGTTCTG CAATCTTCAA
TTGTCATACA GCAAGACTAT ATAATAGCTT TCAAAATAAA

	2721	ATCATAGGCA	GTTCTCATAA	ATGGAATCAT	GTTTGAACAT
		CCTAATTCTG	TTGGCATGGA	GTGCTTTGAC	ATTTTGAAGT
	2801	GTGCATCAGT	GTGAATAACT	GAATTTCCTT	TTCCCAAGGT
		ATTGTGTTGG	ACTCTGGTGA	TGGTGTCAGC	CACACTGTCC
5	2881	CCATCTATGA	AGGATATGCT	CTCCCCCATG	CTATCCTTCG
		TCTCGACCTT	GCTGGGCGTG	ATCTCACTGA	TTACCTCATG
	2961	AAGATCCTGA	CGGAGCGTGG	TTACTCATTC	ACCACAACGG
		CCGAGCGGGA	AATTGTGAGG	GACATGAAGG	AGAAGCTTTC
	3041	CTACATCGCC	CTGGACTATG	ACCAGGAAAT	GGAGACTGCC
10		AAGACCAGCT	CCTCCGTGGA	GAAGAGCTAC	GAGCTTCCTG
	3121	ATGGACAGGT	TATCACCATT	GGTGCTGAGC	GTTTCCGCTG
		CCCTGAGGTC	CTCTTCCAGC	CTTCCTTCAT	AGGAATGGAA
	3201	GCTGCGGGTA	TCCATGAGAC	TACATACAAC	TCCATCATGA
		AGTGCGACGT	GGATATTAGG	AAGGATCTAT	ATGGCAACAT
15	3281	CGTTCTCAGT	GGTGGTACCA	CTATGTTCCC	TGGCATTGCT
		GACAGGATGA	GCAAGGAAGA	TCACTGCCTT	GCTCCTAGCA
	3361	GCATGAAGAT	CAAGGTGGTC	GCCCCTCCTG	AAAGGAAGTA
		CAGTGTCTGG	ATTGGAGGAT	CCATCTTGGC	ATCTCTCAGC
	3441	ACATTCCAGC	AGGTAAATAT	ACAAATGCAG	CAATGTAGTG
20		TTGTTTACCT	CATGAACTTG	ATCAATTTGC	TTACAATGTT
	3521	GCTTGCCGTT	GCAGATGTGG	ATTGCCAAGG	CTGAGTACGA
		CGAGTCTGGC	CCATCCATTG	TGCACAGGAA	ATGCTTCTAA
	3601	TTCTTCGGAC	CCAAGAATGC	TAAGCCAAGA	GGAGCTGTTA
		TCGCCGTCCT	CCTGCTTGTT	TCTCTCTTTT	TGTTGCTGTT
25	3681	TCTTCATTAG	CGTGGACAAA	GTTTTCAACC	GGCCTATCTG
		TTATCATTTT	CTTCTATTCA	AAGACTGTAA	TACCTATTGC
	3761	TACCTGTGGT	TCTCACTTGT	GATTTTGGAC	ACATATGTTC
		GGTTTATTCA	AATTTAATCA	GATGCCTGAT	GAGGGTACCA
	3841	GAAAAAATAC	GTGTTCTGGT	TGTTTTTGAG	TTGCGATTAT
30		TCTATGAAAT	GAATAACATC	GAAGTTATCA	TCCCAGTATT
	3921	TTCGCATGAA	TGTTCTTTTC	TTCTGTCTTG	TGCATCAGTG
		ATCTAGTGCA	TGGGAGTTTG	TATTGTGATG	TTCGACATCA
	4001	CGTAACTTCC	ACTTTGCCTT	TGCTGTTCGA	TATTTTAATG
		ACATGTCACA	CACACTTCTG	CTACTTTTCT	TTCTTGGCTA

4081 TTGTGCCAGC ATGATGCAAG ATGCATCACA CGATCAGATA
TATTCTCATC GTCAGGCTTT CAGGCACAGA GCACGCTTTG
4161 CGCTTAAAAG TTGTACCGCC AGTAGACATC CCCTGTAGAA
GTGATAATCT TTTCACTTTT CTTAAAGAAA TTGAGAGGGG
4241 AAATGGAACC ATGTGGATCA GAGAAGCTTT TGTTTCTTAC
ACAAGAATAT TTGGTACAGT GGGGGTCCTA TGTTCGTGGG
4321 TTCGTGGCTT GGCTCCCTGT CTTCAACCAA GTGTTTTCAG
TTCAACATGT TAGCGTGTAG AAAGAGCACA ATTCTGTTTA
4401 TCTCCAAGGT AAAATGTGGC ATTCTGTTAA AGAACATGAT
CCTGCCAATT TTTTAAGTTT CAATGGAAGA GGAATGTAAA
4481 GCTTTCTATG GTTTGTGTAC ACAACACAGT GGAAGAGGAG
TGCAAGCTTT CTATGGTTTG TGTGCGCGTT GTGGTGCAGC
4561 ACTTCAATTT TGTTAGAAAT GAAAGAAAAA AAAGGATGAT
CATGCTTATA GTAAATCACT CTTTTTCCTC GCCTTCTGTA
4641 CGTTTTGACT TGACAAGATT TTAAAATCTG TACATGACCT
TTGTTTTAAA ATTACTTTAT GTATTTCCAT CTTTCAAGTT
4721 ATGCAGATGT CATCACAAAT TGTTACACCA ATCACCAGGC
TGGCTGTTTA TATATTATCA GACCAGGCTA TATAGAGTAT
4801 ACTATACTAA CTGTTCATAT TATCTGGAAA TCTTGCTTGC
TACTTGAGCG GGAAAAGGGT ATAGATATGA GGTTAAGGAA
4881 CGAAGCGGCA GCAAATCGAG GCTCTCTCTG AAATCATTTT
ACATCTACAA AAGCACATTT AACCTTTTCT AGAACACATA
4961 TGTTACTTAG AAGCAGGAAG TTCATGCAAA ATTTCATCGA
CAAGATAACC AGGCCGGCAC TGGAAGAGTT ATCTTTTACC
5041 TCAATCTGTA TACACTCAAA GTTACTCGGA TTGTACATTG
GCTAAAAGTT TCCCTGTTTC ATTTGAACCA CCTCAGCAAA
5121 AGCAACCTGA AGAGTTTGTT GTGCAAAGGT AAAAACCTTC
CCCCAGACTT TGATCCTTCC CTTGCATATC TAAGGGCATC
5201 ACGGTGAGGT CACTGTACCG CAAGCATTAG TCCAACACAA
AGCCATTCTT TGCTTCTTTT GTCCACCGTT TCAATATGTA
5281 TACATCTGGT ATGGTGCGTA CATCAAGGGC CAAGAATACT
CTTAGTATAT GCCGGCACAA GCTACCACAA CTCTCAAACT
5361 TGCAGCAGCT GCACTTAGCT ATATTGCCAG AAGTATCATA
CCTGACTCTG CATGTGGCTT CAGTATGGTC CTTTGTGACA

5441 CTATACACAG CAATCAACCC ATCATTGTCA AGACTAGAGA
TATATAATAG CCTAAAGATC CAATGAATCC 5510

An alignment between the sequence of the RAc1 cDNA clone and that of the RAc1 genomic sequence was used to determine the structure of the RAc1 gene shown above. Translation of the pcRA1.3 insert in all three reading frames identified a potential coding region of 1131 nucleotides. The potential coding region of the RAc1 cDNA, if translated *in vivo*, would code for 377 amino acids and an actin protein of 41.9 kDA estimated molecular mass. This analysis identified three introns, interrupting the gene at the same places in the RAc1 coding sequence as those previously reported for all other plant actins. The analysis also identified an additional intron within the transcribed sequence of RAc1 which is 5' of the region containing the translation initiation codon. This 5'-intron separated a 79 bp GC-rich 5'-noncoding exon from an exon coding the translation initiation codon.

In the sequence depicted above, the promoter according to the present invention lies within 0 to 2071 nucleotides, more particularly, the efficiency of the promoter appears to lie within about 800 to 2071 nucleotides; nucleotides 811 to 816 and 2066 to 2070 are the XhoI and EcoRV enzyme sites, respectively; 2051 to 3600 nucleotides encompass the coding region DNA sequence for the rice actin gene RAc1; and 1650 to 3841 nucleotides is the RAc1 genomic clone for pRAc1.

To determine if any of the previously isolated plant actin genes also contain 5'-intron-like sequences, the region 5' of their

respective translation initiation codons to that of RAc1, and those animal actin genes known to contain such 5'-introns as compared. This analysis revealed that the soybean (*Glycine max*) actin genes SAc3 and SAc1, the *Arabidopsis thaliana* actin gene AAc1, the potato (*Solanum tuberosum*) actin processed pseudogene PAc-psi, and the maize (*Zea mays*) actin gene MAc1 have regions upstream of their translation initiation codons which bear sequence similarity to the 3'-splice site junction regions of the 5'-introns of the rice actin gene RAc1, a *Xenopus borealis* cytoplasmic actin gene and the *Drosophila melanogaster* cytoplasmic actin gene DmA2. These previously confirmed, and putative intron splice sited are all found within 7 to 11 bp upstream of their respective translation initiation codons. Although the short region of untranslated exons they would encode are all AT-rich, they do not as a group suggest any strong consensus sequence. However, these sequences may serve a similar function in the different actin genes.

The complete structural analysis of the RAc1 gene described above has therefore led to the identification and localization of a 5' noncoding exon, separated by a 5'-intron from the first coding exon, in the RAc1 genomic sequence. It has been reported that a 5'-intron in the maize Adh1 gene is essential for the efficient expression of foreign genes from the maize Adh1 promoter [Callis et al., 'Introns increase gene expression in cultured maize cells. Genes & Development, 1:1183 (1987)]. In order to investigate the effect of the RAc1 5'-intron on gene

expression, a number of GUS fusion plasmids containing RAc1 intron deletions which involved the successive removal of those intron sequence elements previously determined as being important for efficient intron splicing, such as the mRNA branch point and 3'-donor splice sites. The results of transient assays of GUS activity in rice protoplasts transformed with the various deletion constructs suggest that the 5'-intron of RAc1 is essential for efficient gene expression from the RAc1 promoter. Test results also suggest that the intron-mediated stimulation of gene expression is not a function of the intron sequence per se but is associated, in part, with an in vivo requirement for efficient intron splicing.

This analysis and following discussion of the identification and localization of the 5' noncoding exon in the RAc1 genomic sequence can be more easily understood with references to the following examples:

Example II

Genomic Clone Characterization:

A genomic restriction map of the 15.1 kb insert from lambda-RAc1 was prepared by analyzing all possible single and double digests with the enzymes BamHI, EcoRI, HindIII and SalI. A 5.3 kb HindIII-HindIII restriction fragment from the lambda-RAc1 clone was subcloned into pBluescript-KS to generate the plasmid pRAc15'.H3. A restriction map of pRAc15'.H3 was prepared by analyzing all possible single and double digests with the enzymes BamHI, BglII, EcoRI, EcoRV, HincII, HindIII, KpnI, PstI, SmaI, SphI, XbaI and XhoI. Restriction digestion, plasmid ligation, transformation of E. coli DH5-alpha competent cells and isolation of plasmid DNA were done following standard procedures.

A 2.1 kb EcoRV-EcoRV region within the pRAc15'.H3 insert was further subcloned into pBluescript-KS and its DNA sequence determined according to the sequencing strategy outlined in Fig. 2B. Double stranded DNA sequencing reactions were run using the dideoxy chain termination method following T7 polymerase™ protocol. Analysis of DNA sequence data were performed using the Microgenie™ computer program.

EXAMPLE III

Construction of Plasmids:

The 2.1 kb EcoRV-EcoRV fragment from pRAc15'.H3 was subcloned into the SmaI site of pBluescript-KS, in both orientations, to produce the plasmids pRAc15'.21V and pRAc15'.21V^r. A promoterless ß-glucuronidase (GUS) gene, containing the 3' noncoding region of the nopaline synthase (NOS) gene, was excised by BamHI-XbaI digestion of pEXAG3 and cloned between the BamHI and XbaI sites of pRAc15'.21V and pRAc15'.21V^r to produce the plasmids pRAc15'.21VG and pRAc15'.21V^{r}G. Deletion of the 0.8 kb XhoI-XhoI and 1.2 kb EcoRI-EcoRI fragments from pRAc 15'.21VG generated the plasmids pRAc15'.13XG and pRAc15'.09RG respectively. Introduction of the 1.2 kb EcoRI-EcoRI fragment into the EcoRI site of pRAc15'.09RG, in the reverse orientation, produced the plasmid pRAc15'.21VR^{r}G. The plasmid pRAc15'.09RGIDS$^-$ was produced by deletion of a 0,17 kb BamHI-BamHI fragment from pRAc15'.09RG.

The plasmids pRAc15'.09RGIΔ8, pRAc15'.09RGIΔ1 and pRAc15'.09RGIΔ12 were generated by cleavage of pRAc15'.09RG at its BglII site followed by exonuclease III deletion and S1 nuclease treatment to remove different amounts of the RAc1 first intron. To construct the plasmid pRAc15'.09RGI$^-$ the 0.9 kb EcoRI-EcoRV fragment from pRAc15'.H3, containing the first intron of the rice RAc1 gene, was cloned between the EcoRI and EcoRV sites of pBR322 to produce the plasmid pBRAc15'.09R. The 0.4 kb SstI-

EcoRV fragment from pBRAc15'.09R, containing the RAc1 intron, was excised and replaced with the intronless 0.1 kb SstI-EcoRV fragment from the insert of an RAc1 cDNA clone, pcRAc1.3, to produce the plasmid pBRAc15'.09RI$^-$. The intronless 0.6 kb EcoRV fragment from pBRAc15'.09RI$^-$ was excised and cloned between the EcoRI and SmaI sites of pBS-KS to yield the plasmid pRAc15'.09RI$^-$. Cloning of the Gus-Nos containing BamHI-XbaI fragment from pEXAG3 between the BamHI and XbaI site of pRAc15.09RI$^-$ produced the plasmid pRAc15'.09RGI$^-$. The nature of this RAc1 promoter-GUS fusion was confirmed by sequencing double-stranded DNA.

EXAMPLE IV

Culture, transformation and visualization of GUS activity in rice cell suspension cultures and protoplasts:

Cell suspension cultures were generated from calli, initiated from scutella of mature rice (*Oryza sativa* L. v Lemont) seeds, and cultured in liquid R2 media containing 3% (w/v) sucrose, 2 mg/ml 2,4-dichlorophenoxyacetic acid and 2 mg/l vitamin B-5. All cell suspensions were subcultured weekly and incubated in the dark at 26°C. Cell suspensions were filtered through a 700 μm mesh prior to particle gun bombardment with 1.2 μm diameter tungsten particles mixed with a solution containing 10 μg of plasmid DNA. GUS activity in intact cells and developing calli was determined by the appearance of blue spots two and ten days after bombardment with plasmid DNA following the GUS assay procedure of Jefferson et al.

Protoplasts were isolated from rice (*Oryza sativa* L. v Nipponbare) cell suspension cultures and resuspended in MaMg medium to a final density of 10^6 protoplasts/ml. For transformation, 1 ml of protoplast suspension was incubated with 10 μg of circular plasmid DNA, 50 μg/ml of calf thymus carrier DNA and an equal volume of 30% polyethylene glycol 4000. The mixture was incubated for 30 minutes, diluted with CPW13 medium, washed a further 3 times in CPW13 media, with centrifugation between each wash, before being resuspended in simplified KPR liquid media to a final density of 10^6

protoplasts/ml. 0.1 ml aliquots of this final suspension were incubated by thin layer culture in 96-well plates prior to subsequent assays for GUS activity.

EXAMPLE V

Analysis of GUS activity in transformed protoplasts:

To monitor the temporal pattern of GUS expression for each plasmid in transformed rice protoplasts, two wells from each thin layer culture were assayed 1, 3, 5, 10, 20 and 30 days after transformation. Cells were incubated with 5-bromo-4-chloro-3-indolyly glucuronide in a standard assay procedure for 48 hours before visualizing GUS activity.

For the quantitative analysis of GUS activity, cells were collected 20 days after transformation by low speed centrifugation in a bench top microcentrifuge. Total soluble protein was isolated in a GUS extraction buffer. Protein extracts were incubated with 1 mM methylumbelliferyl-ß-D-glucuronide (MUG) in a standard assay at 37°C for 3 to 6 hours. The liberation of 4-methylumbelliferone (MU) was followed by measuring fluorescence with excitation at 365 mM and emission at 455 nM in a spectrofluorometer. Protein concentrations of plant extracts were determined by the dye-binding method of Bradford. Statistical analysis of the quantitative GUS activity results were done using a paired t-test.

To begin the analysis of the rice RAc1 regulatory regions, restriction mapping of a lambda-EMBL4 phage clone, lambda-RAc1, which had previously been shown to contain the rice RAc1 gene was carried out. The resulting restriction map of the 15.1 kb lambda-RAc1 insert is shown in Fig. 2A. This figure also indicates the position of the RAc1 coding and noncoding exons, as previously determined. To isolate and characterize the 5'-flanking sequence of the rice RAc1 gene, a 5.3 kb HindIII fragment from the lambda-RAc1 insert, spanning a region from 3.9 kb upstream of the RAc1 coding sequence to a point within its third coding exon, was isolated and cloned into the HindIII site of the plasmid pBluescript-KS to produce the plasmid pRAc15'.H3. A restriction map of the pRAc15'.H3 insert is shown in Fig. 2B. Restriction maps of the lambda-RAc1 (A) and pRAc15'.H3 (b) inserts were determined by single and double restriction enzyme digestion, as detailed in the methods. The enzyme sites are abbreviated as follows: BamHI, B; BglII, G; EcoRI, E; EcoRV, R; HincII, H2; HindIII, H3; SphI, Sp; SstI S; XhoI, Xh. The position of the 5.3 kb Hind III fragment within the lambda-RAc1 insert is indicated by dashed lines between (A) and (B). Noncoding and coding portions of RAc1 exons are depicted by open and filled boxes respectively. The strategy used to sequence the 2.1 kb EcoRV fragment within the pRAc15'.H3 insert is indicated by horizontal arrows. The length of the horizontal arrows are indicative of the size of the sequence obtained.

The restriction map of the pRAc15'.H3 insert shown in Fig. 2B also indicates the strategy used to sequence an EcoRV fragment which covers a region 2.1 kb upstream of the translation initiation codon of the RAc1 gene. This sequence is as follows:

```
                                                                  T
      801  AGCTAGCATA  CTCGAGGTCA  TTCATATGCT  TGAGAAGAGA   Xho1
           GTCGGGATAG  TCCAAAATAA  AACAAAGGTA  AGATTACCTG
      881  GTCAAAAGTG  AAAACATCAG  TTAAAAGGTG  GTATAAGTAA
           AATATCGGTA  ATAAAAGGTG  GCCCAAAGTG  AAATTTACTC
      961  TTTTCTACTA  TTATAAAAAT  TGAGGATGTT  TTGTCGGTAC
           TTTGATACGT  CATTTTTGTA  TGAATTGGTT  TTTAAGTTTA   i
     1041  TTCGCGATTT  TGGAAATGCA  TATCTGTATT  TGAGTCGGGT   i
           TTTAAGTTCG  TTTGCTTTTG  TAAATACAGA  GGGATTTGTA
     1121  TAAGAAATAT  CTTTAAAAAA  ACCCATATGC  TAATTTGACA
           TAATTTTTGA  GAAAAATATA  TATTCAGGCG  AATTCTCACA   Eco RI
     1201  ATGAACAATA  ATAAGATTAA  AATAGCTTGC  CCCCGTTGCA
           GCGATGGGTA  TTTTTTCTAG  TAAAATAAAA  GATAAACTTA
     1281  GACTCAAAAC  ATTTACAAAA  ACAACCCCTA  AAGTCCTAAA   ii
           GCCCAAAGTG  CTATGCACGA  TCCATAGCAA  GCCCAGCCCA   iii
     1361  ACCCAACCCA  ACCCAACCCA  CCCCAGTGCA  GCCAACTGGC
           AAATAGTCTC  CACACCCCGG  CACTATCACC  GTGAGTTGTC
     1441  CGCACCACCG  CACGTCTCGC  AGCCAAAAAA  AAAAAAAGAA   iv
           AGAAAAAAAA  GAAAAAGAAA  AAACAGCAGG  TGGGTCCGGG
     1521  TCGTGGGGGC  CGGAAAAGCG  AGGAGGATCG  CGAGCAGCGA
           CGAGGCCGGC  CCTCCCTCCG  CTTCCAAAGA  AACGCCCCCC
     1601  ATCGCCACTA  TATACATACC  CCCCCCTCTC  CTCCCATCCC   v
           CCCAACCCTA  CCACCACCAC  CACCACCACC  TCCTCCCCCC   vi
     1681  TCGCTGCCGG  ACGACGAGCT  CCTCCCCCCT  CCCCCTCCGC   Sst I
           CGCCGCCGGT  AACCACCCCG  CCCCTCTCCT  CTTTCTTTCT
     1761  CCGTTTTTTT  TTTCGTCTCG  GTCTCGATCT  TTGGCCTTGG
           TAGTTTGGGT  GGGCGAGAGC  GGCTTCGTCG  CCCAGATCGG
     1841  TGCGCGGGAG  GGGCGGGATC  TCGCGGCTGG  CGTCTCCGGG
           CGTGAGTCGG  CCCGGATCCT  CGCGGGGAAT  GGGGCTCTCG   Bam HI
```

1921 GATGTAGATC TTCTTTCTTT CTTCTTTTTG TGGGTAGAAT BglII
TTGAATCCCT CAGCATTGTT CATCGGTAGT TTTTCTTTTC
2001 ATGATTTGTG ACAAATGCAG CCTCGTCCGG AGCTTTTTTG vii
TAG *GTAGAAG ATG GCT GAC GCC GAG GAT ATC* Eco RV
Met Ala Asp Ala Glu Asp Ile

In this sequence, nucleotides are numbered with the A of the RAc1 transcription initiation site designated in bold. Restriction sites used in the subsequent construction of the various RAc1-GUS fusion plasmids and structural regions described in detail are underlined and named or designated by lower case Roman numerals and noted at the right margin of the sequence table. Upper case capital letters represent 5'-flanking sequences, upper case italic letters represent exon sequences and upper case capital letters between the two sets of exon sequences represent the intron sequence. The codons of the RAc1 first coding exon have their translation product indicated below them.

A number of potential regulatory sequences were identified in the 5'-flanking region of RAc1 gene. A 12 bp direct repeat, GGTTTTAAGTT (region i), is located between bases 1027 to 1038 and 1078 to 1088. A tandem (imperfect) direct repeat of 16 bp, AA(G/C)CCC(T)AAAGT(G/C)CTA (region ii), is located between bases 1301 and 1333. 20 bp downstream of this tandem direct repeat are eight tandem copies of an imperfectly repeating pentamer with the consensus sequence CCCAA (region iii). Finally, between bases 1465 and 1505, there is a purine rich sequence where 35 out of 40 bases are "A"s (region iv)

Downstream of these putative regulatory regions a number of sequences have been identified which have previously been implicated in the control of gene expression. Between bases 1609 and 1617 there is a putative TATA box (region v). The sequence and position of the putative RAc1 TATA box are in agreement with that previously determined for a number of other plant genes, and the sequence around the transcription initiation site; CCTACCA is similar to the consensus sequence for transcription initiation previously determined for a number of animal YAYY and plant YYYAYYA genes (Y= pyrimidine). The noncoding exon located 3' of the TATA box is GC rich (77.5%) and consists of a number of tandemly repeated (A/T)CC triplets (region vi).

The 5'-intron of the rice RAc1 gene is 313 bp long. Its 5'-acceptor splice site [(G/G)TA], and 3'-donor splice site [TTTTTTGTA(G/G)], follow the consensus sequence previously determined for a number of plant genes. A putative branch point site for mRNA splicing between bases 2008 and 2012 (region vii) was identified whose sequence, GTGAC, and distance from the 3'-donor splice site bears similarity to the location and consensus sequence for animal mRNA branch point splice sites, YTRAC.

Four RAc1-GUS fusion plasmids were constructed to determine the minimum amount of RAc1 5'-flanking sequence required for maximal ß-glucuronidase (GUS) gene expression; these are shown in Fig. 3. The plasmid pRAc15'.21VG has the 2.1 kb EcoRV restriction fragment from pRAc15'.H3 fused to the GUS coding region and 3'-noncoding transcript terminator region of the

nopaline synthase (NOS) gene. This plasmid encodes a transcript containing two in frame translation start codons, one each from the RAc1 and GUS genes, adding 15 amino acids to the N-terminal end of the wild type GUS protein. The plasmids pRAc15'.13XG and pRAc15'.09RG were created by the deletion of 0.8 and 1.2 kb of sequence, respectively, from the 5' end of pRAc15'.21VG. The plasmid pRAc15'.21VR^{r}G has the 5' 1.2 kb EcoRI fragment cloned in the opposite orientation to that in pRAc15'.21VG.

A number of plasmids were also constructed to investigate the importance of the RAc1 5'-intron on RAc1-promoter-mediated gene expression. The plasmid pRAc15'.09RGIΔ8 contains a deletion of 9 bp around the BglII site of the RAc1 5'-intron. The plasmid pRAc15',09RGIΔ1 contains a deletion of 133 bp between bases 1868 and 2003 of the sequence depicted above, reducing the distance between the 5' acceptor and mRNA branch point from its 5'-intron. The plasmid pRAc15'.09RGIΔ12 has a deletion of 157 bp between bases 1868 and 2027, removing the putative mRNA branch point from its 5'-intron. The plasmid pRAc15'.09RGIDS^{-} was created by excision of a 170 bp BamHI fragment from pRAc15'.09RG, removing the putative mRNA branch point and 3'-donor splicing sites from the RAc1 5'-intron. The plasmid pRAc15'.09RGI^{-} is identical to pRAc15'.09RG except that it lacks the entire RAc1 5'-intron.

The plasmid pRAc15'.21V^{r}G contains the 2.1 kb EcoRV fragment from pRAc15'.H3 cloned in the opposite orientation to that in pRAc15'.21VG. It was postulated that the RAc1 5'-region

would not show bidirectional promoter activity because the 5'-most 1.2 kb EcoRV-EcoRI restriction fragment in pRAc15.21VG did not show any binding when used as a probe in Northern hybridization against total RNA from 7 day old rice shoots. As a positive control, the plasmid pAI$_1$GusN was used. This plasmid contains the promoter, first exon and first intron of the maize Adh1 gene fused to a GUS coding sequence with the 3' noncoding region of the NOS gene.

More specifically, the individual maps of constructs containing various portions of the 5'-flanking and 5'-transcribed sequence of the rice RAcl gene fused, in frame, to a sequence containing the GUS coding region and 3' NOS transcription terminator are shown in Figure 3. Open boxes represent noncoding exons of the RAc1 gene, filled boxed represent the GUS coding region (not to scale) and striped boxes represent the NOS terminator sequence. The stippled box in the construct pAI$_1$GusN represents the first exon of the maize Adh1 gene. The 'Δ' symbol indicates the deletion point in the various RAc1 first intron deletion constructs. The indicated restriction enzyme sites are abbreviated as follows: BamHI, B1; BglII, B2; EcoRI, R1; EcoRV, RV; HincII, H2; HindIII, H3; SstI, S1; XhoI, Xh1; XbaI, Xb1.

To assay GUS expression from the various RAc1-GUS fusion plasmids, rice suspension culture cells were transformed by the biolistic method [see Wang, Y.C. et al., Transient expression of foreign genes in rice, wheat, and soybean cells following particle bombardment. Plant Mol. Biol. 11:433 (1987); and Cao, J. et al.,

Transformation of rice and maize using the biolistic process. UCLA Symposia on Plant Gene Transfer (1989)], and assayed for high level GUS activity by visual inspection 10 days after transformation according to Jefferson et al. [GUS fusions: ß-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J. 6:3901 (1987)].

The results of these various assays are tabulated in the following table:

TABLE I

	Protoplasts:		Cell Suspension
Names	mean specific Activity nmol/hour/µg	mean GUS Expression %	Visible GUS Activity
pRAc15'.21VG	211.4[1]	8.4	+
pRAc15'.13XG	195.0[1]	7.9	+
pRAc15'.09RG	94.2[2]	5.7	+
pRAc15'.21VR^{r}G	92.5[2]	6.1	+
pRAc15'.09RGIΔ8	95.4[2]	2.8	N.A.
pRAc15'.09RGIΔ1	53.4[3]	1.4	N.A.
pRAc15'.09RGIΔ12	17.0[4]	0.2	N.A.
pRAc15'.09RGIDS$^-$	8.8[5]	0.2	N.A.
pRAc15'.09RGI$^-$	4.7[5]	0.2	-
pRAc15'.21V^{r}G	0.7[5]	0.0	-
Untransformed	0.0[5]	0.0	-
pAl$_1$GusN	31.9[6]	4.2	+

N.A. = not assayed

The results of the rapid assay are shown in the third column of Table 1. Suspension cultures transformed with the plasmids pRAc15'.21VG, pRAc15'.13XG, pRAc15'.09RG, pRAc15'.21VR^{r}G and pAI$_1$GusN were all positive for GUS activity while untransformed cultures or those transformed with pRAc15'.09RGI$^-$ or pRAc15'.21V^{r}G displayed no visible GUS expression.

Table I also shows the results of a determination of GUS expression efficiencies, defined as the percent of intact cells displaying visible GUS activity 10 days after PEG-mediated transformation of rice protoplasts with the RAc1-GUS fusion constructs. The plasmids pRAc15'.21VG and pRAc15'.13XG displayed the highest GUS expression efficiencies at 8.4 and 7.9% respectively. The plasmids pRac15'.09RG and pRAc15'.21VR^{r}G showed GUS expression efficiencies of 5.7 and 6.1%, respectively. All four of the aforementioned plasmids displayed significantly higher GUS expression efficiencies than was found for pAI$_1$GusN. Removal of RAc1 5'-intron sequences resulted in a significant reduction in the number of cells displaying visible GUS activity, relative to pRAc15'.09RG, which contains the intact RAc1 5'-intron. For plasmids pRAc15'.09RGIΔ8 (2.8%) and pRAc15'.09RGIΔ1 (1.4%), the longer the sequence deleted between the 5'-acceptor and the putative mRNA branch point splice sites, the lower the GUS expression efficiency. Removal of the branch point for mRNA splicing in the plasmid pRAc15'.09RGIΔ12 and the additional removal of the 3'-donor splice site in the plasmid pRAc15'.09RGIDS$^-$ led to GUS expression efficiencies which, at

0.2%, represents less than 4% of that determined for the plasmid pRAc15'.09RG. Untransformed protoplasts or protoplasts transformed with either pRAc15'.21V^{r}G or the intronless plasmid pRAc15'.09RGI^{-} gave no blue cells.

To determine the quantitative differences in GUS expression from the various RAc1-GUS fusion plasmids, the constructs were introduced into rice protoplasts by PEG-mediated transformation, and GUS specific activity was assayed 20 days after the transformation procedure. The results of this quantitative analysis of GUS expression are presented in the first column of Table I.

The highest GUS specific activities were recorded for the plasmids pRAc15'.21VG and pRAc15'.13XG, which have the longest RAc1 upstream regions. The plasmid pAl$_1$GusN, which has the GUS gene expresses from the maize Adh1 promoter, displayed less than 17% of the GU specific activity shown by pRAc15'.21VG and pRAc15'.13XG. The plasmids pRAc15'.09RG, with a truncated RAc1 5'-region, and pRAc15'.21VR^{r}G, with an inverted RAc1 5'-region, were found to have GUS specific activities around 48% of that for pRAc15'.21VG. There was no statistically significant difference (mean specific activity with identical superscripted numbers are not statistically different from each other), in specific activity between these two constructs, suggesting that the inversion of the RAc1 5'-end did not influence GUS expression from the truncated RAc15'.09RG construct.

The effects of the various RAc1 intron deletions could be seen when their GUS specific activities were compared to that of their progenitor plasmid, pRAc15'.09RG. The small 9 bp deletion in the RAc1 5'-intron of plasmid pRAc15'.09RGIΔ8 had no effect on the quantitative level of GUS activity. The plasmid pRAc15'.09RGIΔ1, containing a deletion of 133 bp between the 5'-acceptor splice site and the putative mRNA branch point splice site, displayed a GUS specific activity that was less than 57% of that observed for pRAc15'.09RG. The plasmid pRAc15'.09RGIΔ12, which had the putative mRNA branch point site removed, showed a further decrease in GUS specific activity of 17% of that observed for pRAc15'.09RG. The plasmid pRAc15'.09RGIDS$^-$, which lacks the putative mRNA splicing branch site and 3'-splicing donor site of the RAc1 5'-intron, and pRAc15'.09RGI$^-$, which lacks the entire RAc1 5'-intron, displayed no significant GUS specific activities over that observed for untransformed samples. However, it was noted in Table I that while blue cells, indicative of GUS expression, were never observed in untransformed rice protoplasts or protoplasts transformed with the intronless plasmid pRAc15'.09RGI$^-$, GUS expression was able to be visualized in those rice protoplast transformed with the plasmid pRAC15',09RGIDS$^-$.

As a first step towards an understanding of the regulation of rice actin gene expression, the present invention has described the isolation and characterization of the 5'-regions of the rice RAc1 gene. Within the 5'-flanking sequence, the occurrence of a

long poly(dA) stretch located between bases 1465 and 1505 which appears to play a role in the constitutive activation of RAc1 gene expression was found. Based upon chemical analysis, it has been suggested that a minimal affinity for histone cores and nucleosome formation is provided by homogeneous tracts of purines located on one strand of the double helix. Poly(dA) regions within recombinant DNA molecules can prevent nucleosome formation *in vitro*, and it has also been found that naturally occurring poly(dA) tracts act as 5'-promoter elements for the constitutive expression of different yeast genes. It is believed that poly(dA) stretches may change the chromatin structure, enabling general transcription factors to access the DNA template and activate constitutive transcription in the absence of more specific transcription factors.

The sequence of RAc1 untranslated mRNA 5'-region is unusually GC rich, with an AT content of only 26.5%. In a survey of 79 plant genes it was found that their mRNA 5'-sequences were extremely AT rich, with 71 of the genes surveyed having an AT content greater than 51% and only one of the genes surveyed having an AT content of less than 44%. In the ribosomal-scanning model of translation initiation it is proposed that the AT richness of the mRNA 5'-sequence leads to the formation of relatively labile secondary structures which facilitate the movement of the ribosomal subunits towards the translation initiation codon. The significance of the GC rich RAc1 mRNA 5'-sequence, with respect

to transcript stability and translation efficiency, remains to be investigated.

The construction of a number of RAc1-GUS fusion constructs has provided the determination that the plasmid pRAc15'.13XG, containing 0.83 kb of 5'-flanking sequence, the noncoding exon and 5'-intron of the RAc1 gene, has the minimal amount of RAc1 sequence necessary for maximal GUS expression in transient assays of transformed rice protoplasts; the additional 0.8 kb of RAc1 5'-sequence presents in the pRAc15'.13VG did not significantly increase GUS expression above that of the plasmid pRAc15'.13XG. It was also found in the present invention that the RAc1 5'-flanking sequence was more active than the maize Adh1 5'-flanking region in stimulating GUS expression in transformed rice protoplasts. Deletion of a 0.4 kb region from the 5'-end of the pRAc15'.13XG insert resulted in a 52% reduction in GUS activity. However, no statistically significant difference in GUS specific activities was detected between the plasmids pRAC15'.09RG, with the 0.4 kb region deleted, and pRAc15'.13.09RG, with the 0.4 kb deleted, and pRAc15'.21VR^{r}G, with the 0.4 kb region present but inverted and displaced 0.8 kb further upstream. This suggests that the sequence element(s) in the 5'-end of the pRAc15'.13XG insert that are responsible for high level GUS expression do not display any position/orientation-independent, enhancer-like activity. In most cases the differences in GUS specific activities detected between the RAc1-GUS fusion constructs could be directly correlated with

their differential GUS expression efficiencies. This suggests that there is a threshold to the visualization of GUS expression. Protoplasts transformed with the various RAc1-GUS fusion constructs are presumed to display a position-effect-dependent distribution of GUS specific activities, the mean of which is a function of their particular RAc1 promoter. Therefore those protoplast populations transformed with the most active RAc1 promoter-GUS fusion constricts will display the greatest number of visually detectable blue cells and the highest GUS expression efficiencies.

By constructing RAc1-intron-deletion-GUS plasmids, it was able to be shown that GUS expression in transformed rice protoplasts was dependent on the presence of an intact RAc1 5'-intron. Deletion of the RAc1 intron reduced GUS specific activity to levels that were not significantly greater than that of untransformed protoplasts. In transient assays of transformed maize protoplasts a maize Adh1-S gene lacking the nine Adh1-S introns was expressed at levels 50- to 100-times lower than that of the intact gene. The reintroduction of the 5'-intron of the Adh1-S gene was enough to restore the level of expression to that observed for the intact gene, although this effect was only observed if the first intron was reinserted in a 5'-position; it is believed that this was a result of increases in the amount of mature, cytoplasmic mRNA and not a result of increased transcript stability in the nucleus or increased translation efficiency in the cytoplasm.

The increase in expression resulting from the presence of RAc1 5'-intron is not believed to be caused by the presence of some enhancer-like cis acting sequence within the intron. No sequences within the RAc1 intron were found which bore any homology to either an enhancer-like sequence from the first intron of the human ß-cytoplasmic actin gene, or to any sequences within the maize Adh1 5'-intron. Nor is it believed that the RAc1 first intron codes for any functionally active transcript or protein product. Previous Northern blot hybridization with a double stranded probe that spans the RAc1 first intron failed to show any binding to total RNA from seven day old rice shoots, other than to the 1.7 kb RAc1 transcript itself. Finally, the RAc1 5'-intron contains no open reading frames of any significant length. One conclusion of the present invention suggests that the primary effect of RAc1 first intron is associated with an *in vivo* requirement for efficient splicing.

The small deletion in the RAc1 intron in the plasmid pRAc15'.09RGIΔ8 caused no significant reduction in GUS specific activity relative to that of pRAc15'.09RG. However, a significant difference was observed between the GUS specific activities of the plasmids pRAc15'.09RG and pRAc15'.09RGIΔ1. The 133 bp deletion in the plasmid pRAc15'.09RGIΔ1, while not removing any of the sequences previously implicated in intron splicing, did cause a reduction in the distance between the 5'-acceptor and putative mRNA branch point splice sites with an associated 44% reduction in GUS specific activity. Large reductions in splicing

efficiency occur *in vivo* and *in vitro* when the distance between the 5'-acceptor and branch point splice sites is reduced; this requirement for a minimal distance between the 5'-acceptor and branch point splice sites probably reflects a requirement for multiple splicing factors to interact with specific intron regions, these regions being presumably removed in the construction of pRAc15'.09RGIΔ1. The *in vivo* requirement for efficient intron splicing was further supported by the observation that a deletion which removed the putative 5'-intron mRNA branch point splice site in the construction of pRAc15'.09RGIΔ12, or removed both the branch point and 3'-donor splice sites in the construction of pRAc15'.09RIDS⁻ led to GUS specific activities which were 18% and 9% of that observed for pRAc15'.09RG. Deletion of the normal branch point and 3'-donor splice sites of animal genes does not abolish splicing but rather leads to the use of cryptic branch point and 3'-donor splice sites, with associated reductions in splicing efficiency and gene expression; this may be occurring in rice protoplasts transformed with the plasmids pRAc15'.09RGIΔ12 and pRAc15'.09RGIDS⁻, leading to reduced but still detectable GUS expression

The results of the present invention clearly demonstrate that the RAc1 5'-flanking sequence between 0 and 2070 nucleotides and more particularly between 800 and 2070 nucleotides in the genomic sequence contains an efficient promoter for rice transformation. Furthermore, the expression of a foreign gene in transformed plant cells can be dependent upon,

but not necessarily an absolute function of, the presence of an intact 5'-intron sequence. It appears that a functional requirement for the presence of the 5'-intron may be correlated with the conservation of the 5' noncoding exon during this rice actin gene's structural evolution. In addition the results of the present invention indicate that the maize 5'-intron is located between the first and second coding exons of the Adh1 gene, while the rice 5'-intron is located between a 5' noncoding exon and the first coding exon of the RAc1 gene, suggesting that there may be a common positional component to the intron mediated stimulation of gene expression observed for the maize and rice 5'-intron.

In summary, the preceding description of the present invention clearly demonstrates that a 2.1 kbp 5' of the Act1 gene's translation initiation codon, containing 1.3 kb of 5' untranscribed sequence, the 5' transcribed but untranslated exon, 5'-intron and part of the first coding exon of the rice Act1 gene, is capable of conferring high level expression of foreign gene in transformed rice material. Thus this region can be used to activate the constitutive expression of foreign genes in transgenic plants of rice and other agronomically important plants; the 5'-intron of the rice Act1 gene can stimulate the expression of a foreign gene in transformed rice material [thus this (and the other introns of the rice Act1 gene) will be able to increase the expression of foreign genes in transformed plants of rice (and other agronomically important plants) when ligated in front of such a gene in between a promoter and the gene]; and

although no specific function for the 3'-end of the rice Act1 gene in the regulation of foreign gene expression in rice has been demonstrated. It appears from the present invention that the 3'-end of the rice Act1 gene should also stimulate the expression of such foreign genes in transformed plants of rice and other agronomically important crops.

Thus, while we have illustrated and described the preferred embodiment of our invention, it is to be understood that this invention is capable of variation and modification and we therefore do not wish or intend ourselves to be limited to the precise terms set forth, but desire to avail ourselves of such changes and alterations which may be made for adapting the invention to various usages and conditions. Accordingly, such changes and alterations are properly intended to be within the full range of equivalents, and therefore within the purview of the following claims.

Having thus described our invention and the manner and a process of making and using it in such clear, full, concise and exact terms so as to enable any person skilled in the art to which it pertains, or with which it is mostly nearly connected, to make and use the same;

WE CLAIM:

1. A promoter for monocotyledonous plants including a DNA sequence comprising a 5'-region of a rice RAc1 gene that directs the expression of a linked structural gene in the tissues of a recipient plant.

2. A promoter for monocotyledonous plants according to Claim 1 comprising the region extending from approximately position 1 to approximately position 2070 of the RAc1 gene.

3. A promoter for monocotyledonous plants according to Claim 2 containing the region extending from approximately position 800 to approximately position 2070 of the RAc1 gene.

4. A recombinant monocotyledonous plant comprising a structural gene for the production of a plant produced product linked with a promoter comprising a 5'-region of a rice RAc1 gene that directs the expression of the linked structural gene.

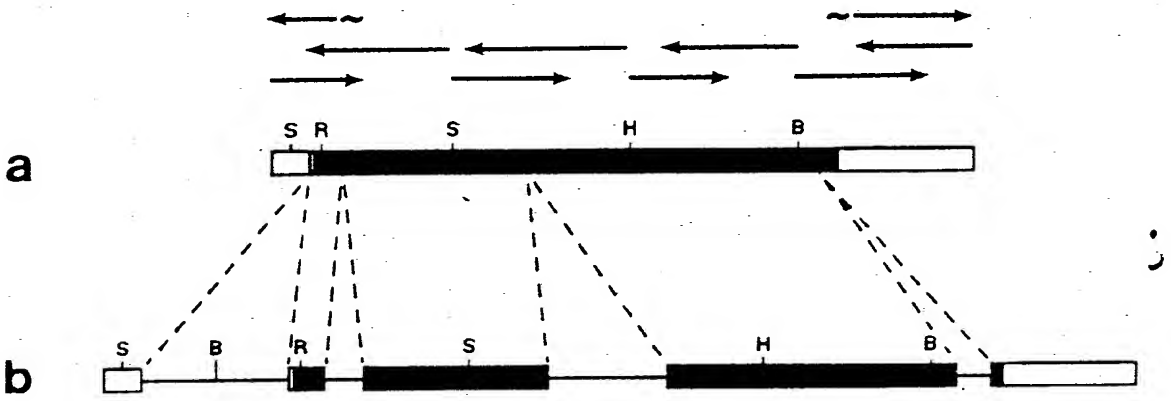
5. A genetic construct selected from the group consisting of pRAc15'.21VG, pRAc15'.13XG, pRAc15'.09RG, pRAc15'.21VR^{r}G, pRAc15'.09RGIΔ8, and pRAc15'.09RGIΔ1.

6. A regulatory element for RAc1 gene promoter activity comprising approximately a 1.3 kb genetic sequence located in the 5'-region upstream portion of the RAc1 translation initiation codon.

7. A genetic construct comprising the RAc15' intron ligated in between a suitable promoter and a structural gene to enhance the expression of the latter.

8. A genomic sequence comprising the RAc1 sequence including a 5' non-coding exon separated by a 5'-intron from the first coding exon in the sequence.

9. A genomic sequence comprising a fragment having a restriction map selected from:



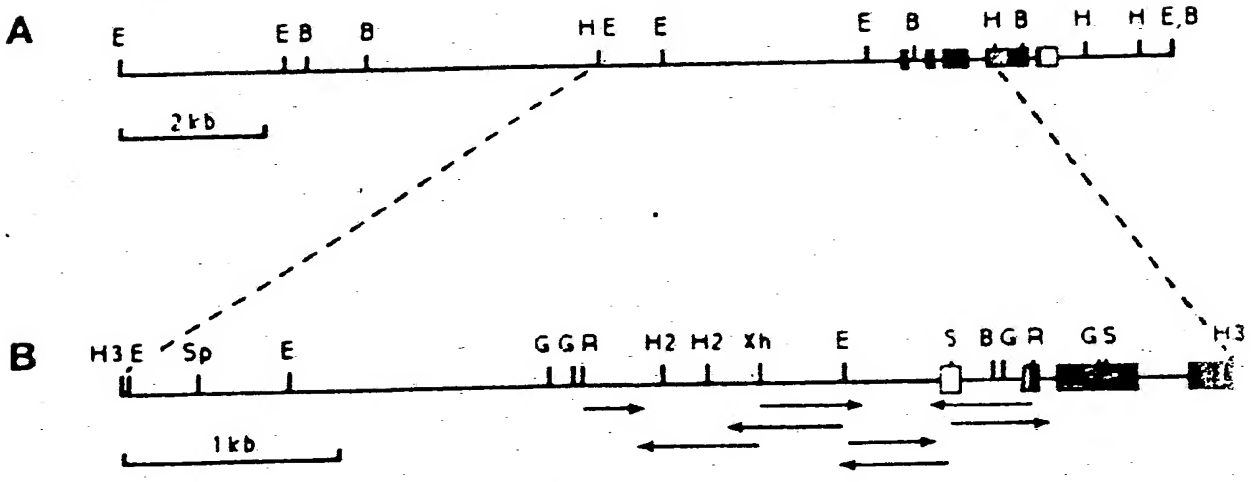


FIG. 1

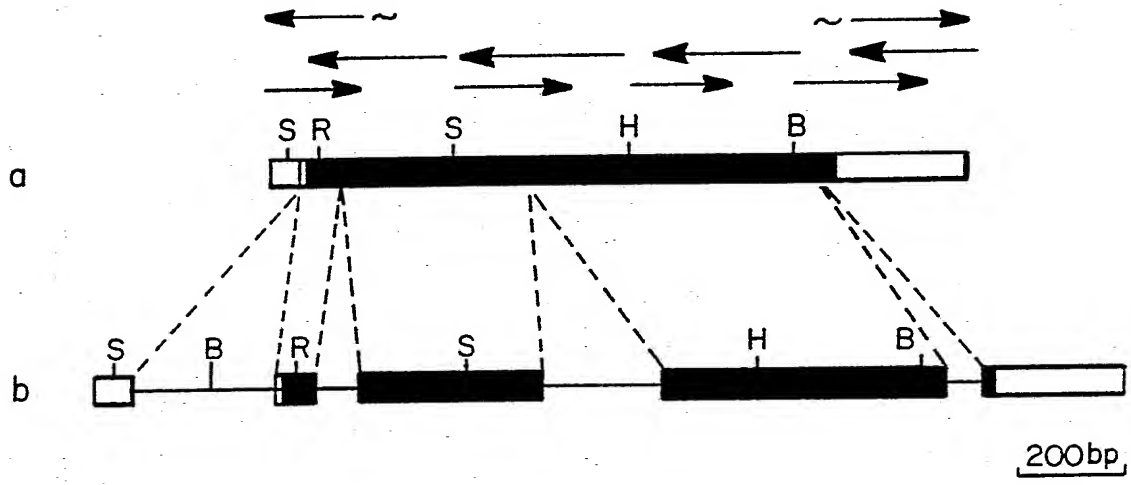
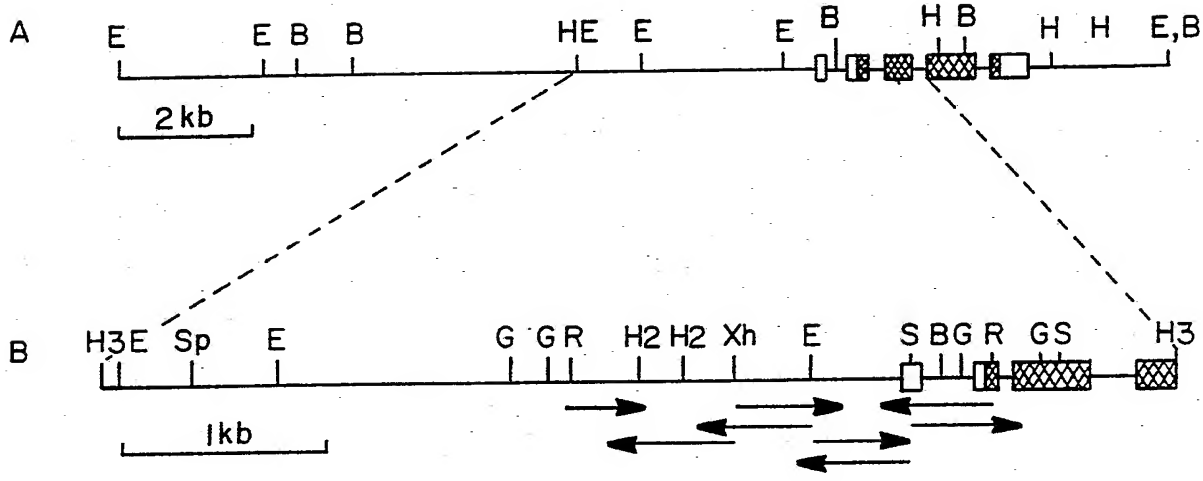
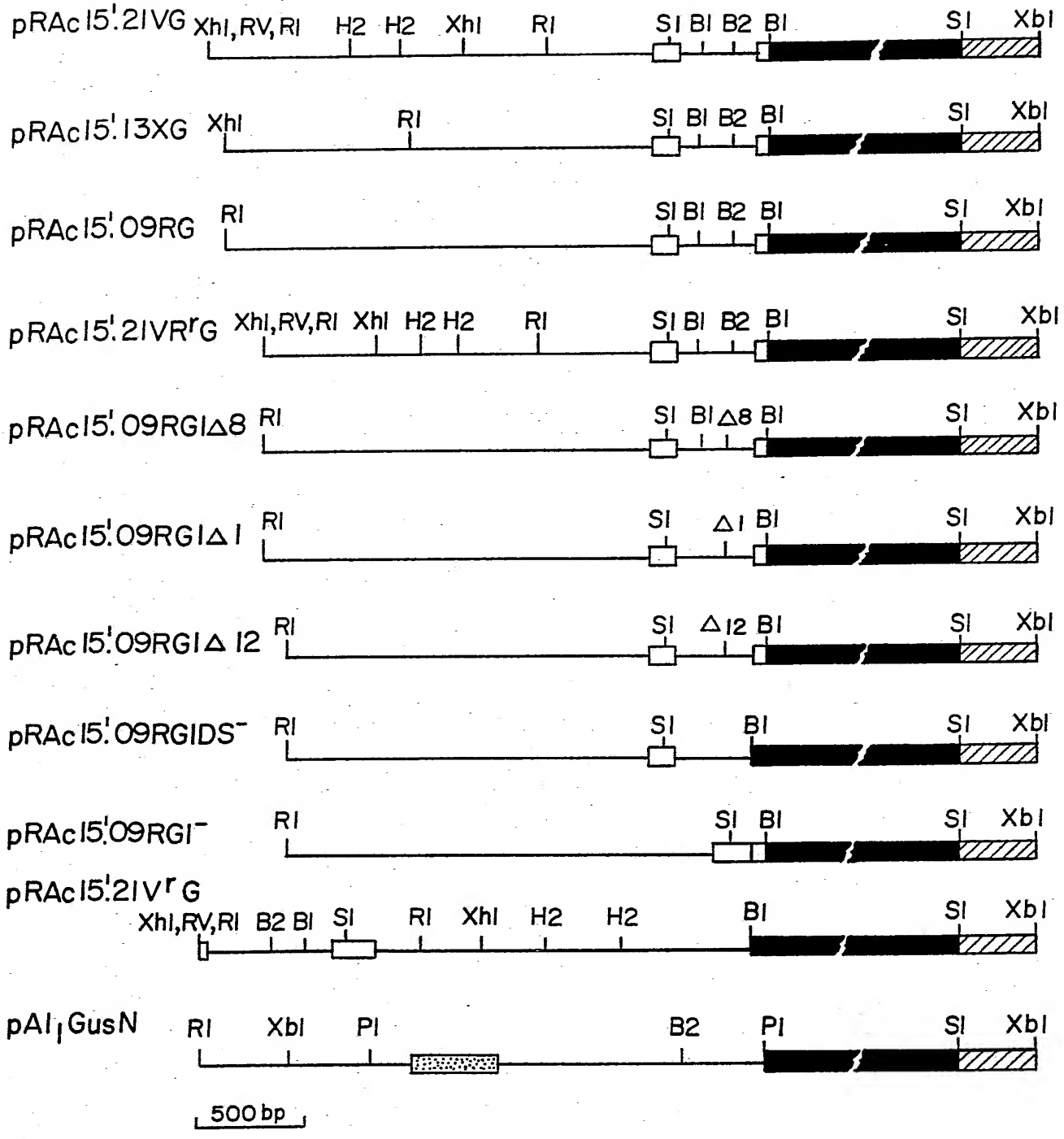

a
b
S R S H B
S B R S H B
200bp

FIG. 2


A
E E B B HE E E B H B H H E,B
2kb
B
H3E Sp E G G R H2 H2 Xh E S B G R G S H3
1kb

2/2

FIG. 3



SUBSTITUTE SHEET